

УДК 534.44; 534.785

*В. В. МИТЯНОК*

**ОПРЕДЕЛЕНИЕ ЧИСЛОВЫХ ХАРАКТЕРИСТИК ВЫСОКОЧАСТОТНЫХ ЗВУКОВ  
РЕЧИ НА ОСНОВЕ АППРОКСИМАЦИИ ГАРМОНИЧЕСКИМИ ФУНКЦИЯМИ**

*Полесский государственный университет, Пинск*

*(Поступила в редакцию 26.10.2007)*

Несмотря на многочисленные попытки, задача уверенного распознавания компьютером человеческой речи до сих пор не решена. Не решена также задача распознавания компьютером человека по его голосу, хотя известно, что хорошо знакомые между собой люди легко узнают друг друга при разговоре по телефону.

Тем не менее, несмотря на то, что в целом проблема еще не решена, имеется ряд несомненных достижений. Полученные к настоящему времени результаты и еще нерешенные проблемы весьма полно изложены в [1].

В современных системах распознавания человеческой речи используются разные идеи. Так, первоначально еще Гельмгольц полагал, что задачу распознавания речи можно решить, если изучить механизмы речеобразования – строение гортани, носовых пазух, полости рта, выделить в них резонаторы, исследовать возникающие в них стоячие волны и др. В настоящее время чаще всего изучаются наборы различных признаков, характеризующих звуки, с использованием статистического анализа [2]. Используются уровни в спектральных полосах, формантные признаки, коэффициенты линейного прогноза [2–3]. В [3] выдвинуто предположение об определяющей роли соотношения уровней мощности в определенных спектральных полосах речевого сигнала. Но ни один из возможных наборов не дает явного преимущества по сравнению с любым другим, поэтому некоторые из них применяют одновременно. Так, в [4] для идентификации звуков используются параллельно формантные и полосные признаки. Подробное освещение вопроса можно найти в [5].

Во многих системах распознавания речи требуется раздельное произнесение слов, с паузой между ними не менее 150 мс [6]. Лишь год назад появились системы NaturallySpeaking фирмы Dragon Systems и ViaVoice фирмы IBM, не требующие таких пауз, но этот успех был достигнут не за счет улучшенной математической проработки задачи, а за счет совершенствования компьютеров. Программы на первом этапе анализируют звук, чтобы отличить низкочастотные гласные от высокочастотных согласных. После этого результаты сравнивают с фонемами, и по результатам сравнения происходит распознавание звука. Эти программы являются дикторозависимыми, т. е. перед первым использованием программа должна «привыкнуть» к голосу диктора [7].

Надо полагать, что основная сложность распознавания звуков связана с недостаточной математической проработкой задачи, желанием добиться скорейшего финансового успеха, что приводит к тому, что многие, существенные для распознавания звуков, математические особенности звуковых кривых до сих пор остаются неизученными. В этой связи особенно интригующим выглядит отмеченное в [8] обстоятельство, что «речевой сигнал достаточно хорошо воспринимается человеком даже в очень узкой полосе частот, причем расположенной в любой части речевого диапазона. Это свойство совершенно не соответствует механизмам обработки речи, принятым в системах автоматического распознавания». Совершенно очевидно, что без ответа на вопрос, почему это происходит (а также на ряд других), дальнейший прогресс в решении проблемы распознавания звуков невозможен.

Эти обстоятельства стимулируют проведение дальнейших исследований для выяснения тех *математических* особенностей звуковых кривых, которые существенны для распознавания как речи, так и говорящего. Задача настоящей статьи – изложение метода определения числовых характеристик звуков, основанного на решении задачи аппроксимации в противовес широко используемому преобразованию Фурье.

Большинство существующих систем распознавания человеческой речи основано на разложении звуков в спектр преобразованием Фурье [9–11]. Применительно к распознаванию реальных звуков здесь существуют две серьезные проблемы: первая – определение длительности отрезка звуковой кривой, отобранного для изучения; вторая – нестрогая периодичность звуковой кривой даже при произнесении долгих гласных звуков. Так, известно, что даже идеальная гармоническая кривая (сигнал)

$$y(t) = \sin(\omega_0 t), \quad (1)$$

будучи подвергнутой Фурье-преобразованию на ограниченном интервале изменения независимой переменной (времени)  $t$ , приводит к «размазанному» спектру, содержащему кроме главного, также и бесконечно большое количество побочных (фальшивых), т. е. не соответствующих наличию излучения с соответствующей частотой максимумов.

В самом деле, синус-преобразование функции (1) на симметричном интервале времени

$$C_s(\omega) = \int_{-t_0}^{t_0} y(t) \sin(\omega t) dt \quad (2)$$

приводит к результату

$$C_s(\omega) = \frac{\sin((\omega_0 - \omega)t_0)}{(\omega_0 - \omega)} - \frac{\sin((\omega_0 + \omega)t_0)}{(\omega_0 + \omega)}. \quad (3)$$

(Косинус-преобразование здесь дает тождественный нуль.) Анализируя выражение (3), нетрудно заметить, что зависимость  $C_s(\omega)$  содержит бесконечное число локальных максимумов, главный максимум имеет место при  $\omega = \omega_0$ , остальные же являются побочными. По мере увеличения  $t_0$  побочные максимумы сгущаются к главному максимуму. В частности, в квантовой механике это обстоятельство является математической подоплекой хорошо известного соотношения неопределенностей [12].

С побочными максимумами можно бороться путем увеличения длительности сигнала, однако применительно к реальному звуку это приводит к усугублению проблемы нестрогой периодичности. Дрейф частоты, нестабильность амплитуд практически не сказываются на спектре коротких сигналов, но, по мере роста их длительности, вносят все большие искажения в спектр.

Проиллюстрируем это обстоятельство графически на одном частном, но показательном примере. Пусть преобразованию Фурье подвергается сигнал с качающейся несущей частотой, описываемый выражением

$$y_i = \sin(i(0,1 + 0,01 \sin(1 + i \cdot 0,002))) \quad (4)$$

Качания частоты сигнала (4) происходят вокруг ее среднего значения 0,1 с амплитудой 0,01 и частотой качаний, равной 0,002. Пусть длительность сигнала составляет 2000 точек ( $I = 1..2000$ ).

Преобразование Фурье сигнала (4) приводит к спектру, график которого показан на рис. 1.

Данный график не позволяет распознать в сигнале наличие всего лишь одной, пусть даже и качающейся частоты. Увеличение же длительности сигнала (4) приводит лишь к тому, что количество максимумов на спектральной кривой возрастает, они становятся все более резкими, и разобратся в этой картине становится все более затруднительным.

Все это означает, что при разложении в спектр преобразованием Фурье, отрезок реальной звуковой кривой, подлежащей анализу, должен быть с одной стороны, не слишком коротким, а с другой – не слишком длинным. То есть должен быть найден компромисс по длительности. Но, каким бы ни было компромиссное решение о выборе длительности звучания, распознаванию звуков будут мешать как слишком малая, так и слишком большая его длительность.

Отмеченные выше трудности являются характерными для преобразований Фурье. Но их можно избежать при ином подходе к проблеме. А именно – вместо разложения сигнала в спектр преобразованием Фурье следует решать задачу аппроксимации. При таком подходе длительность сигнала вообще не имеет значения, а нестабильность параметров мод вносит лишь небольшие искажения в спектр. Рассмотрим этот подход подробнее.

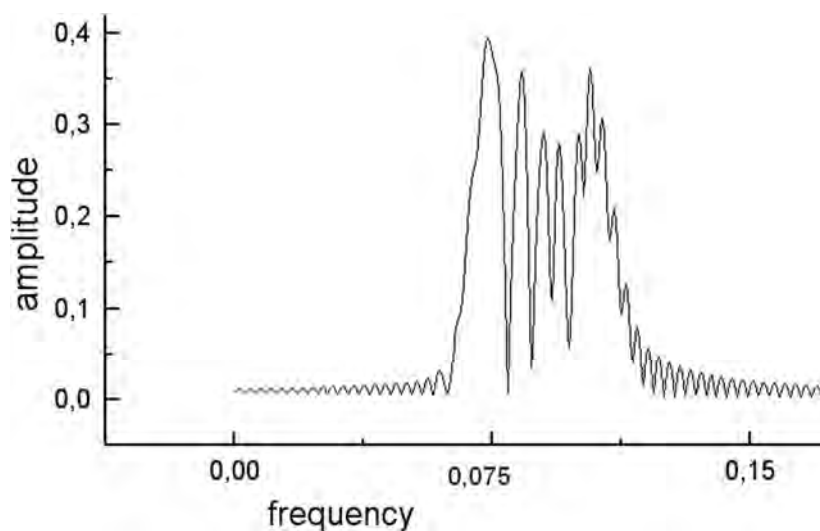


Рис. 1. Спектр сигнала с качающейся несущей частотой, полученный преобразованием Фурье

Пусть требуется аппроксимировать функцию  $y_i = y(x_i)$ , заданную своими значениями в избранных точках  $x_i$  с помощью гармонических функций с медленно изменяющимися амплитудами. Для упрощения положим  $x_i = i$ . С целью аппроксимации сначала рассмотрим невязку (функционал)

$$S = \sum_{i=1}^n (y_i - b_0 - \sum_{k=1}^{l_1} a_k \sin(\omega_k i) - \sum_{k=1}^{l_1} b_k \cos(\omega_k i))^2, \quad (5)$$

где  $l_1$  – количество гармоник (мод),  $\omega_k$  – их частоты (считаются известными),  $n$  – количество используемых точек кривой звукового давления,  $a_k, b_k$  – соответственно амплитуды синус- и косинус-волн,  $b_0$  – нуль звуковой кривой. Индекс  $i$  нумерует те значения аргумента звуковой кривой, в которые было произведено измерение звукового давления. Минимизация (5) по  $a_k, b_0, b_k$ , где  $k = 1 \dots l_1$ , и последующее решение получившейся системы линейных алгебраических уравнений приводит к решению задачи аппроксимации с постоянными, не зависящими от номера точек, амплитудами. В частности, если аппроксимируемая функция  $y_i$  является суммой нескольких гармоник с постоянными амплитудами, частотами и фазами, то (5) достигает теоретического минимума, равного нулю, при совпадении частот, амплитуд и фаз аппроксимируемой и аппроксимирующих функций, причем *независимо от длительности сигнала*. Если же при такой аппроксимируемой кривой частоты аппроксимирующих функций известны с некоторой ошибкой, то, конечно, после решения задачи аппроксимации и подстановки решения в (5) остаточная невязка нулю не равна, но, осуществляя подбор (дрейф) частот, можно найти истинные частоты как частоты, обеспечивающие остаточную невязку, равную нулю.

Такой подход не оправдан в тех случаях, когда по каким-то причинам параметры аппроксимируемой функции медленно (и тем более, не медленно) меняются во времени. Это может случиться, например, в задачах распознавания реальных звуков, произнесенных человеком, хотя бы потому, что в силу физиологических причин голос человека слегка дрожит. Кроме того, при переходе от звука к звуку можно предвидеть изменение как значений частот, так и их амплитуд, фаз. Если звуковую кривую обрабатывать по вышеописанной методике, то следует ограничиваться лишь фрагментами звуковых кривых небольшой длины, так как при большой длине рассмотрения вызванные дрожаниями искажения накапливаются и в итоге приводят к значительным погрешностям в спектре звука.

Однако можно предвидеть, что при распознавании человеческой речи придется иметь дело и с достаточно длительными сигналами. Поэтому для решения вышеозначенной проблемы прием, что амплитуды мод могут медленно (по сравнению с частотами  $\omega_k$ ) меняться со временем. Соответственно, вместо (5) рассмотрим невязку в виде

$$S = \sum_{i=1}^n (y_i - b_{0i} - \sum_{k=1}^{l_1} a_{ki} \sin(\omega_k i) - \sum_{k=1}^{l_1} b_{ki} \cos(\omega_k i))^2 + \alpha \sum_{i=1}^{n-1} (b_{0,i} - b_{0,i+1})^2 + \alpha \sum_{k=1}^{l_1} \sum_{i=1}^{n-1} (b_{k,i} - b_{k,i+1})^2 + \alpha \sum_{k=1}^{l_1} \sum_{i=1}^{n-1} (a_{k,i} - a_{k,i+1})^2. \quad (6)$$

В отличие от (5), в (6) амплитудам и нулю звуковой кривой разрешено меняться от точки к точке, соответственно у них появляется второй индекс, указывающий на номер точки на кривой звука. Однако, чтобы эти изменения не были слишком резкими, в невязку вводится группа членов, учитывающая скачки амплитуд и дрейфующего нуля от точки к точке. Неотрицательное число  $\alpha$  показывает относительную важность различных слагаемых. Если  $\alpha = 0$ , то можно добиться идеальной аппроксимации звуковой кривой, правда, ценой того, что амплитуды и нуль аппроксимирующей функции могут оказаться резко меняющимися от точки к точке. Если же  $\alpha = \infty$ , то амплитудам запрещено изменяться, что фактически эквивалентно (5) и что во многих случаях не соответствует физическому содержанию задачи. Истина, как правило, находится где-то посередине. Поэтому  $\alpha$  будем выбирать так, чтобы вклад в остаточную невязку (6) от членов, содержащих и не содержащих  $\alpha$ , был примерно одинаков. Несколько забегая вперед, отметим, что для этого лучше всего выбрать  $\alpha = 10$ , но варьирование  $\alpha$  в интервале (1...100) практически не сказывается на окончательных результатах исследований.

Предположим, что набор частот  $\omega_k$  известен. Для нахождения минимума (6) требуется приравнять нулю производную  $S$  по каждой из независимых переменных  $a_{ij}$   $b_{ij}$ . Таким образом, общее число уравнений равно

$$n_1 = (2 * l_1 + 1) * n. \quad (7)$$

Дифференцирование по  $b_{0,i}$  приводит к уравнениям

$$b_{0i}(1 + 2\alpha) - \alpha b_{0,i-1} - \alpha b_{0,i+1} + \sum_{i=1}^{l_1} a_{ki} \sin(\omega_k i) + \sum_{k=1}^{l_1} b_{ki} \cos(\omega_k i) = y_i \quad i = 1 \dots n; \quad (8)$$

дифференцирование по  $a_{m,l}$  – к уравнениям

$$\sum_{k=1}^{l_1} a_{k,l} \sin(\omega_k l) \sin(\omega_m l) + \sum_{k=1}^{l_1} b_{k,l} \cos(\omega_k l) \sin(\omega_m l) + b_{0,l} \sin(\omega_m l) + 2\alpha a_{m,l} - \alpha a_{m,l-1} - \alpha a_{m,l+1} = y_l \sin(\omega_m l), \quad (9)$$

где  $m = 1 \dots l_1, l = 1 \dots n$ ;

дифференцирование по  $b_{m,l}$  – к уравнениям

$$\sum_{k=1}^{l_1} a_{k,l} \sin(\omega_k l) \cos(\omega_m l) + \sum_{k=1}^{l_1} b_{k,l} \cos(\omega_k l) \cos(\omega_m l) + b_{0,l} \cos(\omega_m l) + 2\alpha b_{m,l} - \alpha b_{m,l-1} - \alpha b_{m,l+1} = y_l \cos(\omega_m l), \quad (10)$$

где  $m = 1 \dots l_1, l = 1 \dots n$ .

Система уравнений (8) – (10) – суть система линейных алгебраических уравнений относительно амплитуд мод и дрейфующего нуля и методы решений таких систем хорошо известны. Что касается частот мод, то они, как правило, заранее неизвестны. Поэтому их первоначально можно задать генератором случайных чисел и затем осуществить движение (дрейф) методом скорейшего спуска вплоть до достижения наименьшего из всех минимумов остаточной невязки.

Для изучения реальных звуков по вышеописанной методике были собраны образцы звучания звуков С, СЬ, Ш, ШЬ, Ж, ЖЬ, З, ЗЬ, полученные от 4-х мужчин и 4-х женщин.

Запись звуков производилась при частоте дискретизации, равной 192 кГц. Записанные звуки разрезались на образцы длиной в 250 отсчетов. (Рассматривались и другие варианты выбора длин образцов. Оказалось, что образцы длиной менее, чем 200 отсчетов, позволяли распознавать звук

с намного меньшей надежностью, а обработка образцов с длиной более, чем 300 отсчетов, требовала чрезмерно больших затрат машинного времени без существенного улучшения надежности.) Всего от каждого из респондентов для каждого из звуков было получено по 5000 образцов, каждый из которых изучался независимо от других.

Для определения количества мод была принята следующая схема. Количество мод ( $l_1$ ) принималось равным 3 или 4, кроме того, учитывался дрейфующий нуль. Затем, по вышеописанной схеме, математически проводилось уточнение частоты (численная подстройка). Продрейфованные частоты от каждого набора из 5000 фрагментов сводились воедино, и по результатам формировалась гистограмма частот. Оказалось, что каждый раз гистограмма имела от 6 до 13 четко выраженных максимумов. Именно таким и является число мод для каждого звука каждого из респондентов. Соответствующие этим максимумам значения частот и являются искомыми. Однако этот набор частот меняется как при смене респондента, так и при смене звука. То есть, можно сказать, что обнаруженные численные значения частот респондента являются его голосовым портретом (табл. 1). Все численные значения частот даются в долях от частоты

$$\omega_0 = \frac{192000}{2\pi} \approx 30\,558 \text{ Гц, условно принятой за единицу.}$$

Таблица 1. Набор частот мод при произнесении звука С

Респондент	$\omega_1$	$\omega_2$	$\omega_3$	$\omega_4$	$\omega_5$	$\omega_6$	$\omega_7$	$\omega_8$	$\omega_9$	$\omega_{10}$	$\omega_{11}$	$\omega_{12}$	$\omega_{13}$
1	0,041	0,083	0,098	0,140	0,168	0,187	0,210	0,240	0,262	0,287	0,352	0,378	0,437
2	0,052	0,073	0,095	0,125	0,147	0,169	0,243	0,438					
3	0,072	0,105	0,133	0,212	0,257	0,313	0,367	0,425	0,558				
4	0,048	0,091	0,120	0,166	0,181	0,220	0,258	0,290	0,315	0,355	0,420		
5	0,068	0,101	0,139	0,177	0,215	0,266	0,310	0,347	0,440				
6	0,057	0,090	0,115	0,155	0,188	0,221	0,261	0,382	0,411				
7	0,052	0,088	0,117	0,129	0,181	0,240	0,293	0,333					
8	0,044	0,081	0,127	0,166	0,192	0,268	0,298	0,349	0,404	0,445			

Объяснить превращение 3–4 мод в 6–13 пиков на гистограмме можно следующим образом. Реальный звук содержит от 6 до 13 мод, каждая из которых может временно ослабевать (рис. 2). Так как метод аппроксимации ориентирован на минимум невязки (6), то для каждого образца звука (длиной в 250 отсчетов), после окончания дрейфа частоты выходят на те их значения, которые соответствуют наиболее громким именно для данного образца модам. А так как в разных образцах наиболее громкими являются разные моды, то это и приводит к тому, что на гистограмме появляется от 6 до 13 истинных максимумов. После формирования гистограмм проводилось повторное расщепление звуков, но уже на базе истинных частот.

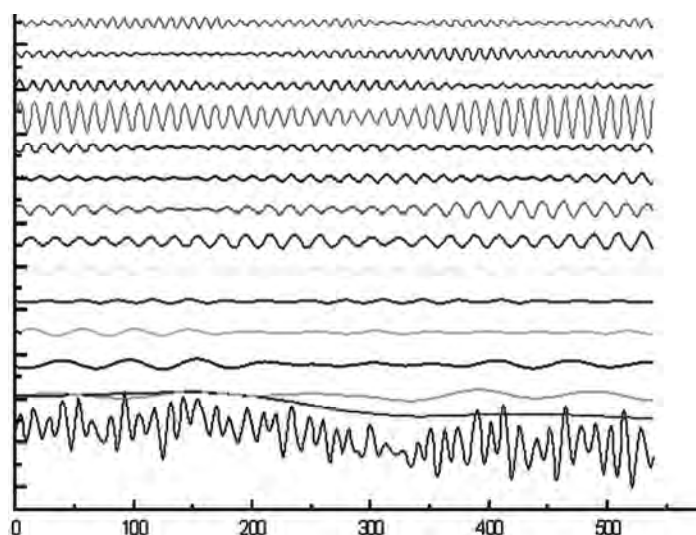


Рис. 2. График звучания звука С и его разложение по модам. Нижний график представляет исходный звук С. Непосредственно над ним – дрейфующий нуль, затем 13 мод по мере увеличения их частот. Две последние моды даны с 10-кратным увеличением

Как видно из рис. 2, отдельные слагаемые звука как бы модулированы по амплитуде, обладают свойством ослабевать, а то и вообще замирать на время.

Так как произнесение звука и его распознавание – это разные процедуры, пусть даже и выполняемые одним и тем же человеком, и заранее неизвестно, кто именно из респондентов будет говорить, то для решения задачи распознавания речи не следует привязываться к какому-то из вариантов базовой системы частот, указанной в табл. 1. Базовый набор для распознавания должен быть универсальным, общим для всех респондентов. Поэтому следует, во-первых, выбрать систему базовых частот и, во-вторых, при избранной системе базовых частот набрать статистический материал по громкости каждой из мод, и тем самым создать эталонную базу данных. Распознавание же звуков можно производить путем сравнения наблюдаемых мод с эталонами базы данных.

Было проанализировано 60 вариантов выбора базовых систем частот. Ниже приведем результаты, полученные по 4-м лучшим системам. Надежность распознавания, как правило, превосходила 50%, но никогда не достигала 100%. (Так как изучались 8 звуков, то вероятность их случайного отгадывания составляла 12,5%.)

Возникла также идея – с целью повышения надежности распознавания заранее провести усреднение спектров по нескольким образцам звучания. Результаты надежности распознавания представлены в табл. 2–5. (В первом столбце таблиц приведена надежность распознавания по одному образцу спектра звучания, во втором – при усреднении по 10 случайно отобраным образцам, в третьем – при усреднении по 100 случайно отобраным образцам.) Под надежностью распознавания понималось точное узнавание звука. Так, например, если звучал звук С, а критерий показывал на близкий звук СЬ, то это считалось ошибкой.

Таблица 2. Надежность распознавания звуков (в процентах) для критерия N1

Респондент	1	10	100
1	50,4	75,8	89,5
2	37,8	49,3	60,7
3	46,7	72,7	87,1
4	49,0	70,4	85,2
5	37,3	48,1	54,3
6	43,9	61,9	71,8
7	38,2	55,1	61,9
8	41,0	50,8	57,9

Примечание. Набор 22 базовых частот 0.105; 0.13; 0.16; 0.185; 0.21; 0.235; 0.26; 0.31; 0.34; 0.37; 0.395; 0.42; 0.45; 0.475; 0.5; 0.525; 0.55; 0.58; 0.63; 0.68; 0.73; 0.79.

Таблица 3. Надежность распознавания звуков (в процентах) для критерия N2

Респондент	1	10	100
1	60,8	89,2	94,5
2	43,3	53,4	57,6
3	51,5	74,8	85,1
4	60,8	78,8	90,7
5	42,4	54,2	65,9
6	50,5	64,2	70,3
7	40,6	57,1	65,5
8	46,5	54,6	59,3

Примечание. Набор 26 базовых частот 0.05; 0.075; 0.1; 0.125; 0.155; 0.18; 0.205; 0.23; 0.26; 0.29; 0.315; 0.345; 0.37; 0.395; 0.42; 0.445; 0.475; 0.5; 0.525; 0.55; 0.575; 0.6; 0.63; 0.68; 0.73; 0.79.

Т а б л и ц а 4. Надежность распознавания звуков (в процентах) для критерия N3

Респондент	1	10	100
1	57,8	85,0	94,5
2	36,5	48,3	57,9
3	43,7	67,6	81,9
4	49,1	70,6	83,7
5	35,3	47,7	56,6
6	37,7	53,5	64,5
7	38,0	48,3	56,8
8	40,4	53,4	66,9

П р и м е ч а н и е. Набор 9 базовых частот 0.052; 0.073; 0.095; 0.125; 0.147; 0.168; 0.192; 0.243; 0.305.

Т а б л и ц а 5. Надежность распознавания звуков (в процентах) для критерия N4

Респондент	1	10	100
1	58,6	80,5	90,3
2	39,4	50,6	60,4
3	47,9	68,7	83,5
4	52,7	63,2	71,6
5	36,4	40,7	48,6
6	45,5	61,3	69,8
7	39,2	46,4	53,2
8	39,3	45,7	55,3

П р и м е ч а н и е. Набор 7 базовых частот 0.072; 0.105; 0.133; 0.212; 0.257; 0.313; 0.367.

Анализируя данные, приведенные в таблицах, можно сделать следующие выводы. 1. Усреднение по нескольким образцам звучания действительно повышает надежность распознавания: чем больше образцов взято для усреднения, тем выше надежность. 2. Надежность распознавания зависит от индивидуальных особенностей респондента, в первую очередь от внятности его речи. 3. Увеличение числа базовых частот свыше 7–9 не приводит к существенному повышению надежности распознавания.

Все это вместе взятое позволяет сделать вывод: *не существует критериев, позволяющих однозначно распознавать произнесенные звуки. Человек распознает звуки по вероятности.*

Так как было обнаружено несколько независимых критериев распознавания звуков, то возникла идея их совместного использования с целью повышения их надежности. Однако оказалось, что механическое объединение каких-то наборов базовых частот в новый базовый набор не приводит к заметному улучшению надежности распознавания, а в ряде случаев приводит к ухудшению. Вероятно, причиной этому является то, что модам становится слишком «тесно». Напротив, раздельное и одновременное использование нескольких критериев дает существенное улучшение (табл. 6).

Как видно из таблицы, надежность распознавания при использовании нескольких критериев заметно возросла по сравнению с надежностью каждого из критериев по отдельности. Не исключено также, что существуют еще какие-то критерии, комплексное использование которых приведет к еще большей степени надежности.

Таблица 6. Надежность распознавания звуков (в процентах) при совместном использовании критериев N3 и N4

Респондент	1	10	100
1	77,9	90,5	96,3
2	69,0	76,3	80,4
3	57,1	74,2	93,5
4	70,3	83,8	91,6
5	52,9	66,1	84,0
6	62,7	82,2	89,3
7	49,5	67,8	79,4
8	59,5	75,5	88,4

### Литература

1. Галунов В. И., Лобанов Б. М., Загоруйко Н. Г. Синтез и распознавание речи. Труды XV сессии Российского акустического общества. М., 2004. С. 3–8.
2. Kuhl P. K., Iverson P. Linguistic experience and the «perceptual magnet effect». In W. Strange. Speech perception and linguistic experiment. Edinburg, 1995. P. 121–154.
3. Галунов В. И., Гарбарук В. И. Акустическая теория речеобразования и системы фонетических признаков. Материалы международной конференции «100 лет экспериментальной фонетике в России». СПб, 2001. С. 58–62.
4. Kraft D. Speech perception // J. Phonetics. 1994. Vol. 7. P. 279–312.
5. www.auditech.ru; www.sakrament.com; www.speechpro.ru
6. Елинек Ф. Разработка экспериментального устройства, распознающего отдельно произнесенные слова // ТИИЭР. 1985. Т. 73, N 11. С. 91–99.
7. [http://www.digest.univers.cv.ua/cnp\\_start.html](http://www.digest.univers.cv.ua/cnp_start.html)
8. Галунов В. И. Современные речевые технологии. <http://g-klishin.narod.ru/works.html>
9. Зигмунд А. Тригонометрические ряды. М., 1965.
10. Бари Н. К. Тригонометрические ряды. М., 1961.
11. Розанов Ю. Н. Спектральные случайные процессы. М., 1964.
12. Ландау Л. Д., Лифшиц Е. М. Квантовая механика. М., 1974.

V. V. MITSIANOK

### DETERMINATION OF NUMERICAL CHARACTERISTICS OF HIGH-FREQUENCY SOUNDS OF SPEECH BY THE HARMONIC FUNCTION APPROXIMATION

#### Summary

The approximation problem-based method of finding the frequencies and movable amplitudes of almost periodical signals is proposed. This method is applied for 8 high-frequency sounds of human speech. The basic mode frequencies for 8 respondents and for 8 high-frequency sounds is established, unique for every person.